

CLAIMS

1. A method for generating a summary of a plurality of related documents in a collection comprising:
 - 5 extracting phrases having focus elements from the plurality of documents;
 - 10 performing phrase intersection analysis on the extracted phrases to generate a phrase intersection table;
 - 15 performing temporal processing on the phrases in the phrase intersection table; and
 - 20 performing sentence generation using the phrases in the phrase intersection table.
2. The method of generating a summary as defined by claim 1, wherein the phrase intersection analysis comprises:
 - 15 representing the phrases in tree structures having root nodes and children nodes;
 - 20 selecting those tree structures with verb root nodes;
 - 25 comparing the selected root nodes to the other root nodes to identify identical nodes;
 - 30 applying paraphrasing rules to non-identical root nodes to determine if non identical nodes are equivalent; and
 - 35 evaluating the children nodes of those tree structures where the parent nodes are identical or equivalent.
3. The method of claim 2, wherein the tree structure is a DSYNT tree structure.
- 25 4. The method of claim 2, wherein the paraphrasing rules are selected from the group consisting of ordering of sentence components, main clause versus a relative clause, different syntactic categories, change in grammatical features, omission of an

empty head, transformation of one part of speech to another, and semantically related words.

5. The method of claim 1, wherein the temporal processing includes:

time stamping phrases based on a first occurrence of the phrase in the
5 collection;

substituting date certain references for ambiguous temporal references;
ordering the phrases based on the time stamp; and
inserting a temporal marker if a temporal gap between phrases exceeds
a threshold value.

10 6. The method of claim 1, further comprising a phrase divergence processing
operation.

7. The method of claim 1, wherein the sentence generation includes mapping
phrases to an input format of a language generation engine and operating the language
generation engine.

15 8. A system for generating a summary of a plurality of related documents in a
collection comprising:

a storage device for storing the documents in the collection;
a lexical database; and
a processing subsystem, the processing subsystem being operatively

20 coupled to the storage device and the lexical database, the processing subsystem being
programmed to access the documents in the storage device and:

using the lexical database to extract phrases having focus elements
from the plurality of documents;

25 performing phrase intersection analysis on the extracted phrases to
generate a phrase intersection table;

performing temporal processing on the phrases in the phrase
intersection table; and

PCT/US00/04118

performing sentence generation using the phrases in the phrase intersection table.

9. The system for generating a summary as defined by claim 9, wherein the phrase intersection analysis processing further comprises:

5 representing the phrases as data structures having root nodes and children nodes;

selecting those data structures with verb root nodes;

comparing the selected root nodes to the other root nodes to identify identical nodes;

10 applying paraphrasing rules to non-identical root nodes to determine if non identical nodes are equivalent; and

evaluating the children nodes of those tree structures where the parent nodes are identical or equivalent.

15 10. The system of claim 9, wherein the data structure is a DSYNT tree structure.

11. The system of claim 9, wherein the paraphrasing rules are selected from the group consisting of ordering of sentence components, main clause versus a relative clause, different syntactic categories, change in grammatical features, omission of an empty head, transformation of one part of speech to another, and semantically related words.

20 12. The system of claim 8, wherein the temporal processing includes:

time stamping phrases based on a first occurrence of the phrase in the collection;

substituting date certain references for ambiguous temporal references;

25 ordering the phrases based on the time stamp; and

inserting a temporal marker if a temporal gap between phrases exceeds a threshold value.

13. The system of claim 8, further comprising a phrase divergence processing operation.

14. The system of claim 8, wherein the processing subsystem includes a language generation engine and wherein sentence generation includes mapping phrases to an input format of the language generation engine and then operating the language generation engine.

5

15. The system of claim 8, wherein the storage device for storing the documents in the collection is remotely located from the processing subsystem.

16. A computer readable media for programming a computer system to perform a method of generating a summary of a plurality of related documents in a collection comprising:

10

extracting phrases having focus elements from the plurality of documents;

15

performing phrase intersection analysis on the extracted phrases to generate a phrase intersection table;

performing temporal processing on the phrases in the phrase intersection table; and

20

25

17. The computer readable media of claim 16, wherein the phrase intersection analysis comprises:

representing the phrases in tree structures having root nodes and children nodes;

selecting those tree structures with verb root nodes;

comparing the selected root nodes to the other root nodes to identify identical nodes;

applying paraphrasing rules to non-identical root nodes to determine if non identical nodes are equivalent; and

evaluating the children nodes of those tree structures where the parent nodes are identical or equivalent.

18. The computer readable media of claim 17, wherein the tree structure is a
5 DSYNT tree structure.

19. The computer readable media of claim 17, wherein the paraphrasing rules are selected from the group consisting of ordering of sentence components, main clause versus a relative clause, different syntactic categories, change in grammatical features, omission of an empty head, transformation of one part of speech to another, and
10 semantically related words.

20. The computer readable media of claim 16, wherein the temporal processing includes:

time stamping phrases based on a first occurrence of the phrase in the collection;

15 substituting date certain references for ambiguous temporal references; ordering the phrases based on the time stamp; and
inserting a temporal marker if a temporal gap between phrases exceeds a threshold value.

21. The computer readable media of claim 16, further comprising a phrase divergence processing operation.

22. The computer readable media of claim 16, wherein the sentence generation includes mapping phrases to an input format of a language generation engine and operating the language generation engine.

007245-124004